TEXT MINING FOR GLOBAL TECHNOLOGY WATCH

- DR. RONALD N. KOSTOFF
- OFFICE OF NAVAL RESEARCH
- CODE 03

- ONR-INDUSTRY DAY
- 5 AUGUST 2004

OUTLINE

- WHAT IS TEXT MINING?
- WHY IS IT IMPORTANT?
- WHAT ARE SOME SPECIFIC BENEFITS?
- HIGH-RISK HIGH-PAYOFF APPLICATIONS!

WHAT IS TEXT MINING

- EXTRACTION OF INFORMATION FROM LARGE VOLUMES OF TEXT
  - COMPUTATIONAL LINGUISTICS
    - QUANTIFICATION OF TEXT ELEMENTS
    - PHRASE FREQUENCIES
    - PHRASE PROXIMITIES
    - DOCUMENT CLUSTERING
  - BIBLIOMETRICS
    - PERFORMERS
    - CENTERS OF EXCELLENCE

WHY IS TEXT MINING IMPORTANT

- GLOBAL S&T ~$1 TRILLION/YR
- ORDERS OF MAGNITUDE GREATER THAN ANY AGENCY'S OR COMPANY S&T
- IMPORTANT TO MAINTAIN AWARENESS OF GLOBAL S&T
  - AVOID TECHNOLOGICAL SURPRISE
  - AVOID REDUNDANCY
  - ALLOWS COORDINATION, LEVERAGING, EXPLOITATION

WHAT ARE SPECIFIC BENEFITS: COMPUTATIONAL LINGUISTICS

- ENHANCING INFORMATION RETRIEVAL AND INCREASING AWARENESS OF THE GLOBAL TECHNICAL LITERATURE

- POTENTIAL DISCOVERY AND INNOVATION BASED ON MERGING COMMON LINKAGES AMONG VERY DISPARATE LITERATURES

- UNCOVERING UNEXPECTED ASYMMETRIES FROM THE TECHNICAL LITERATURE

- ESTIMATING GLOBAL LEVELS OF EFFORT IN S&T SUB-DISCIPLINES

- TRACKING MYRIAD RESEARCH IMPACTS ACROSS TIME AND APPLICATIONS AREAS

WHAT ARE SPECIFIC BENEFITS: <u>BIBLIOMETRICS</u>

- IDENTIFY THE INFRASTRUCTURE (AUTHORS, JOURNALS, INSTITUTIONS) OF A TECHNICAL DOMAIN
- IDENTIFY EXPERTS FOR INNOVATION-ENHANCING TECHNICAL WORKSHOPS AND REVIEW PANELS,
- DEVELOP SITE VISITATION STRATEGIES FOR ASSESSMENT OF PROLIFIC ORGANIZATIONS GLOBALLY,
- IDENTIFY IMPACTS (LITERATURE CITATIONS) OF INDIVIDUALS, RESEARCH UNITS, ORGANIZATIONS, AND COUNTRIES
- GENERATE EXTENSIVE BACKGROUND MATERIAL FOR RESEARCH PAPERS, AND FOR COMPREHENSIVE LITERATURE REVIEWS AND SURVEYS.

LITERATURE-BASED DISCOVERY (LBD)

- REVOLUTIONARY TECHNOLOGIES
- DISRUPTIVE TECHNOLOGIES
- HIGH-RISK HIGH-PAYOFF TECHNOLOGIES
- NOVEL TECHNOLOGIES
- OUT-OF-THE-BOX THINKING
- LATERAL THINKING
- INNOVATIVE INITIATIVES
- DYNAMIC INITIATIVES

LBD (CONT'D)

- MODERN S&T VERY SPECIALIZED
- RESEARCHERS VERY SPECIALIZED
- MANY LITERATURES DISJOINT
- SOME PROBLEMS REQUIRE SOLUTION FROM DISPARATE LITERATURES
- NEED SYSTEMATIC APPROACH TO INTEGRATE FINDINGS FROM DISPARATE LITERATURES

LBD-CONT'D

- LITERATURE-BASED DISCOVERY OFFERS POTENTIAL TO INTEGRATE DIAPARATE LITERATURES

- USES INTERMEDIATE LITERATURES TO RELATE DISPARATE LITERATURES

- COULD BE USED AS STAND-ALONE DISCOVERY GENERATOR

- COULD BE USED AS PRECURSOR TO WORKSHOP

LBD-EXAMPLES

INITIAL APPLICATION-RAYNAUD'S DISEASE (RD)

RETRIEVED RD LITERATURE

IDENTIFIED MAIN FEATURES OF RD

– RHEOLOGY, CLOTTING, VASOCONSTRICTION, ETC

RETRIEVED TOTAL RHEOLOGY LITERATURE

EXAMINED NON-RD RECORDS

IDENTIFIED FISH OIL, EICOSAPENTANOIC ACID AS POSSIBLE TREATMENTS FOR RD

VERIFIED BY CLINICAL/ LABORATORY TESTS

LBD EXAMPLES CONT'D

- LATER APPLICATION-BIOWARFARE AGENT PREDICTION
- IDENTIFIED TWO MAJOR CHARACTERISTICS OF BIOWARFARE AGENTS
  - LETHALITY/ PATHOGENICITY
  - SURVIVABILITY/ TRANSMISSIBILITY
- RETRIEVED LITERATURE FOR EACH CHARACTERISTIC
- FOUND LITERATURES DISJOINT
- IDENTIFIED VIRUSES ADDRESSED IN BOTH LITERATURES
- SOME EXISTING AGENTS; OTHERS POTENTIAL NEW AGENTS

SUMMARY

- GLOBAL S&T AWARENESS IMPORTANT

- TEXT MINING PROVIDES AWARENESS BASED ON DOCUMENTATION

- LITERATURE-BASED DISCOVERY COULD OFFER BASIS FOR NEW RESEARCH DIRECTIONS